Image Fusion in the Medical Field

Digital Image Processing

For the past few decades, there has been a rapid development of image information fusion techniques, particularly in the medical field. Medical image fusion is really important in aiding clinical diagnosis by combining complementary information from different imaging modalities. These modalities include Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Single-Photon Emission Computed Tomography (SPECT).

Each imaging technique provides unique and valuable information about the human body.
- CT images provide detailed structural information, particularly for bones and hard tissues.
- MRI excels in capturing soft tissue details and structural characteristics.
- PET and SPECT are functional imaging modalities that reveal metabolic activity, blood flow, and other physiological processes.

The motivation behind medical image fusion is to combine the strengths of these different techniques into a single comprehensive view. By fusing complementary information from multiple sources, clinicians can gain a more complete understanding of the patient's condition, aiding diagnosis and treatment planning.

There are three key requirements for an effective medical image fusion method:

1. The fused image should retain all relevant information from the source images without any loss of critical details.
2. The fusion process should not introduce any synthetic artifacts, distortions, or false information into the fused image.
3. Potential issues like mis-registration (spatial misalignment) and noise should be avoided or minimized in the fused output.

Fusion Methods:

Here we look at three main domains of medical image fusion techniques: the spatial domain methods, transform domain methods, and deep learning-based methods.

- Spatial Domain Methods:

These early fusion methods operate directly on the image pixels, applying various rules and transformations to combine the source images. The spatial domain techniques include:
a) High-pass filtering: Separates high-frequency (detail) and low-frequency (smooth) components of an image for selective fusion.

b) Principal Component Analysis (PCA): A statistical technique that transforms the input images into a new set of uncorrelated components, enabling fusion by retaining the most significant components.

c) Intensity-Hue-Saturation (IHS) transform: A color space transformation that separates the intensity (brightness) component from the color information (hue and saturation). This helps preserve the color details from functional imaging modalities like PET/SPECT when fused with structural MRI data. IHS method is widely used for fusing functional PET/SPECT images (which provide color information) with structural MRI data. The intensity component from the PET/SPECT image is typically extracted, transformed (e.g., using Log-Gabor filters), and fused with the MRI data. The fused intensity component is then combined with the original hue and saturation components to reconstruct the fused color image.

d) Averaging: A simple fusion rule that computes the pixel-wise average of the input images.

e) Maximum/minimum selection: Selects the maximum or minimum pixel value from the input images at each location.

f) Brovey transform: A technique that preserves the relative spectral contributions of each input image in the fused output.

These spatial domain methods though often suffer from spectral and spatial distortions in the fused output, leading to a shift in research focus towards transform domain techniques.

- Transform Domain Methods:
These state-of-the-art fusion methods work in transformed domains, such as frequency or multi-scale representations, rather than directly on the image pixels. There are three key transformations:

a) Non-subsampled Contourlet Transform (NSCT):
- Provides a multi-scale and multi-directional decomposition of the input images.

- Overcomes the translation-invariance issues present in the basic contourlet transform.
- The source images are decomposed into low-frequency (coarse) and high-frequency (detail) subbands, representing different scales and directional information.
- Various fusion rules are applied to combine the corresponding low and high-frequency subbands from the input images.
- Fusion rules for low-frequency subbands include regional energy-based weighting, sparse representations, and coupled neural networks like PCNN (Pulse-Coupled Neural Network).
- Fusion rules for high-frequency subbands involve techniques like maximum selection, local energy-based weighting, and modified PCNN models.
- After fusing the subbands, the inverse NSCT is applied to reconstruct the fused image.

b) Non-subsampled Shearlet Transform (NSST):
- Provides multi-scale and multi-directional properties similar to NSCT but with lower computational complexity.
- Overcomes the limitations of having a fixed number of directional components present in NSCT.
- The NSST decomposition is often combined with PCNN-based fusion rules for the subbands.

c) Discrete Wavelet Transform (DWT):
- One of the earliest and widely used transform domain techniques for medical image fusion.
- Decomposes the input images into approximate (low-frequency) and detail (high-frequency) coefficients, representing different scales and orientations.
- Various fusion rules are applied to combine the corresponding approximate and detail coefficients from the input images.
- Fusion rules include simple averaging, fuzzy c-means clustering, and complex wavelet extensions like dual-tree complex wavelet transform (DTCWT).
- After fusing the coefficients, the inverse DWT is applied to reconstruct the fused image.

There are some hybrid algorithms too that combine different techniques, such as cascading NSCT with DWT, or incorporating optimization methods like particle swarm optimization (PSO) or predator-prey optimizer (PPO) to improve the fusion process.

- Deep Learning Methods:

With the recent surge in deep learning, medical image fusion research has also explored the use of deep neural networks for end-to-end fusion. The two main extensively used deep learning approaches are:
a) Convolutional Neural Networks (CNNs):
- CNNs learn to fuse images in an end-to-end manner by training on datasets of paired input and ground truth fused images.
- Architectures like Siamese networks and fully convolutional networks have been employed for medical image fusion tasks.
- One of the main challenges here though is the limited availability of large annotated medical imaging datasets for training these data-hungry deep models.
- There are also some other issues, such as: handling convergence issues, overfitting, and optimizing the network architectures for fusion tasks.

b) U-Nets:
- The U-Net architecture, originally developed for medical image segmentation, has been adapted for fusion tasks.
- It follows an encoder-decoder structure with skip connections, allowing the network to preserve spatial details while fusing semantic information.
- U-Nets have been successful in addressing semantic conflicts and mitigating detail loss in the fused output, a common issue with traditional fusion methods.
- Techniques like bilinear interpolation and automatic encoders have been combined with U-Nets for improved fusion performance.
- Research on U-Nets for medical image fusion is still relatively new, but presents several opportunities for further exploration.

Digital Subtraction Angiography (**DSA**):
- It is another technique to visualize vessels in various parts of the body. It particularly proves to be valuable for diagnosing conditions affecting blood flow, such as aneurysms, stenosis or vascular malformations.
- DSA involves capturing X-ray images of blood vessels and then digital subtraction, in the spatial domain.
- Some deep learning models though, particularly convolutional neural network (CNNs), can be trained to denoise DSA images too. This complement support by deep learning methods can help in enhancing image quality, automating analysis tasks, and improving diagnostic outcomes.

Modalities are often fused into pairs in order to get better results. Some commonly **fused modality pairs** are:

1. MRI and PET:

- Fusion of MRI (structural information) and PET (functional information) is crucial for tumor diagnosis, detection of Alzheimer's disease, and brain tumor studies.
- PET images provide metabolic and functional details, while MRI contributes anatomical context and structural details.
- The fusion aids in localizing lesions, identifying tumor boundaries, and understanding the relationship between metabolic activity and brain structures.

2. MRI and CT:
- Combining MRI (soft tissue details) and CT (bone and hard tissue details) provides a comprehensive view of anatomical structures.
- The fusion helps in applications like surgical planning, radiation therapy, and musculoskeletal analysis.
- Research efforts have focused on preserving the complementary information from both modalities while enhancing contrast and structural similarity in the fused output.

3. MRI and SPECT:
- Fusing MRI (structural) and SPECT (functional) data aids in localizing lesions, detecting vertebral bone metastasis, and studying conditions like tinnitus.
- SPECT provides functional information, while MRI contributes anatomical context and soft tissue details.

4. MRI and Ultrasound:
- The fusion of MRI and ultrasound images is also very valuable for applications like vascular imaging and blood flow analysis.
- MRI provides high-resolution anatomical details, while ultrasound captures real-time functional information related to blood flow and tissue motion.

Specific **quantitative metrics** are used for evaluating the quality of fused medical images. These metrics are:

1. Entropy (EN):
- Measures the information content or richness of details in the fused image.
- Higher entropy values typically indicate better fusion performance, as more information from the source images is preserved.

$$\mathrm{EN} = -\sum_{L=0}^{L-1} p_i \times \log_2 p_i,$$

where L represents the number of gray levels and is a probability density function for each gray value i. The entropy is proportional to the amount of information contained in the fused image.

2. Standard Deviation (STD):
- The standard deviation is mainly used to measure the overall contrast of the fused image and is used to determine the difference between the data and the average.
- If the STD value is larger, the more useful information the fused image contains, the better the fusion effect performs, and the image is clearer.

$$\mathrm{STD} = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(f(i,j)-\mu)^2}{MN}},$$

where M and N represent the length and width of the image f(i,j) which is generally 256. The average value of the fused image is represented by μ.

3. Mutual Information (MI):
- Quantifies the degree of similarity or correlation between the fused image and the input source images.
- Higher mutual information suggests that the fused image effectively captures information from all input modalities.

$$\mathrm{MI} = I(x, f) + I(y, f),$$

$$I(x, y) = \sum_{y \in Y}\sum_{x \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)},$$

where p(x) and p(y) are the edge probability density functions of the two images, and p(x,y) is the joint probability density function of the fused image and the source image X, Y.

4. Peak Signal to Noise Ratio (PSNR):
- PSNR is a quantitative measurement method based on mean square error.
- In the fusion image, the higher the PSNR is, the better the SNR is, and the closer it is to the source image.

$$PSNR = 10 * \log_{10}\left(\frac{L^2}{RMSE^2}\right).$$

It represents the maximum pixel gray value in the fused image, which is generally 255. RMSE is the mean square error, and its calculation formula is:

$$RMSE = \sqrt{\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}[\text{ground}(m,n) - \text{fused}(m,n)]^2}{M \times N}}.$$

Mean square error is an image quality measurement method. The value of RMSE is inversely proportional to the quality. The lower the value of RMSE, the better the quality of the fused image have.
ground(m,n) and fused(m,n) represent the intensity values of the source image and the fused image pixel, respectively, and the length and width of the image are M and N, respectively.

5. Structural Similarity Index Measure (SSIM):
- Assesses the structural similarity between the fused image and the source images, considering aspects like luminance, contrast, and structure.
- Higher SSIM values indicate better preservation of structural details from the input modalities.

$$SSIM_{(A,B,F)} = 0.5 \times \left(SSIM_{(A,F)} + SSIM_{(B,F)}\right).$$

In

$$SSIM_{(A,F)} = \frac{(2\mu_A\mu_F + C_1)(2\sigma_{AF} + C_2)}{\left(\mu_A^2 + \mu_F^2 + C_1\right)\left(\sigma_A^2 + \sigma_F^2 + C_2\right)},$$

$$SSIM_{(B,F)} = \frac{(2\mu_B\mu_F + C_1)(2\sigma_{BF} + C_2)}{\left(\mu_B^2 + \mu_F^2 + C_1\right)\left(\sigma_B^2 + \sigma_F^2 + C_2\right)}.$$

where μ_A, μ_B, and μ_F are the average values of the source image and the fused image, respectively; $\sigma^2{}_A$, $\sigma^2{}_B$, and $\sigma^2{}_F$ are the variances of the source image and the fused image, respectively; σAF and σBF represent the joint variance of the two source images and the fused image, respectively.

6. Spatial Frequency (SF):
- The spatial frequency reflects the sharpness of the fused image, that is, the rate of change of the gray image.
- The larger the SF, the higher the image resolution perform.

$$\mathrm{SF} = \sqrt{\mathrm{RF}^2 + \mathrm{CF}^2}.$$

In

$$\mathrm{RF} = \sqrt{\frac{1}{M(N-1)} \sum_{i=1}^{M} \sum_{j=2}^{N} (X(i, j-1) - X(i, j))^2},$$

$$\mathrm{CF} = \sqrt{\frac{1}{(M-1)N} \sum_{i=2}^{M} \sum_{j=1}^{N} (X(i, j) - X(i-1, j))^2}.$$

where RF and CF are the row and column frequencies of the image respectively.

7. $Q^{AB/F}$ Measurement:

- $Q^{AB/F}$ measures the amount of edge information from the source image to the fused image through the Sobel edge detection operator.

- The larger the value of $Q^{AB/F}$ represent, the more information is converted from the source image, and the edge information is better preserved. In general, high edge strength has a greater impact on $Q^{AB/F}$ than low edge strength.

$$Q^{AB/F} = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} \left(Q^A(n, m) W^A(n, m) + Q^B(n, m) W^B(n, m) \right)}{\sum_{n=1}^{N} \sum_{m=1}^{M} \left(W^A(i, j) + W^B(i, j) \right)},$$

where Q^A (n,m), Q^B (n,m) is the edge information storage value; W^A(n,m), W^B(n,m) is the weighting map.

Visual Assessment:
- In addition to quantitative metrics, qualitative visual assessment by human experts (e.g., clinicians, radiologists) is crucial for evaluating the clinical utility and interpretability of the fused images.

- Factors like artefact suppression, detail preservation, and color fidelity are considered during visual evaluation.

Challenges and Future Directions:

There are several key challenges and potential future research directions in the field of medical image fusion:

1. Handling Noise and Artifacts:
- Noise and artifacts present in the source images can be amplified or introduced during the fusion process, degrading the quality of the fused output.
- Developing robust denoising techniques and artifact suppression methods is a critical challenge, particularly for transform domain and deep learning-based fusion approaches.

2. Preserving Colors and Intricate Details:
- Retaining the intricate details, fine structures, and accurate color information from the input modalities is crucial for clinical interpretation and diagnosis.
- Traditional fusion methods often struggle with preserving these nuanced features, leading to blurred or distorted outputs.
- Advanced techniques that can effectively capture and fuse subtle details and color nuances are needed.

3. Optimizing Deep Learning Models and Architectures:
- Deep learning methods for medical image fusion show promising results but face challenges in training data scarcity, model optimization, and architecture design.
- Generating large annotated datasets for training deep models is time-consuming and resource-intensive.
- Developing efficient architectures tailored for the fusion task, while addressing convergence issues and overfitting, is an ongoing area of research.

4. Hybrid Algorithms and Ensemble Methods:
- Combining the strengths of different fusion techniques, such as spatial domain, transform domain, and deep learning methods, through hybrid or ensemble approaches is a promising research direction.
- Hybrid algorithms can leverage the complementary advantages of various fusion strategies, potentially leading to improved performance and robustness.

5. Interpretability and Explainability:

- As deep learning models become more prevalent in medical image fusion, ensuring their interpretability and explainability becomes crucial for clinical acceptance and trust.
- Developing techniques to provide insights into the decision-making process of these models is an active area of research.

6. Real-time and Computational Efficiency:
- Certain clinical applications may require real-time or near real-time fusion capabilities, posing challenges in computational efficiency and hardware acceleration.
- Optimizing fusion algorithms for efficient execution on specialized hardware (e.g., GPUs, FPGAs) is an important consideration, especially for deep learning-based approaches.

****

Reference:

A Review of Multimodal Medical Image Fusion Techniques
Bing Huang, Feng Yang, Mengxiao Yin, Xiaoying Mo, Cheng Zhong, "A Review of Multimodal Medical Image Fusion Techniques", Computational and Mathematical Methods in Medicine, vol. 2020, Article ID 8279342, 16 pages, 2020. https://doi.org/10.1155/2020/8279342